Honours Readings

Matthew Kwan

September 7, 2013

Contents

1 Combinatorial estimates by the switching method

M. Hasheminezhad and B. D. McKay, Combinatorial estimates by the switching method, Contemporary Mathematics, 531 (2010) 209-221.

1.1 Summary

Consider a finite set Ω of objects. A switching is a (nondeterministic) operation that transforms one object into another (or more generally, a switching is a relation $R \subseteq \Omega \times \Omega$). We can partition Ω into subsets $\{C(v)\}_{v \in V}$ and put a directed graph structure (possibly with loops) on the index set V: if an element in $C(v)$ can switch to an element in $C(w)$ then there is an arc between v and w.

Hopefully, for each v we can find a good lower bound $a(v)$ on the number of ways an $\omega \in C(v)$ can be switched, and a good upper bound $b(v)$ on the number of switchings an $\omega \in C(v)$ can be produced by. If we imagine that all switchings are performed at once, a and b give bounds on the inflow and outflow at each vertex in terms of the size of each $C(v)$. (So, $a(v)$ and $b(v)$ can more generally be bounds on the *average* number of switchings per element in $C(v)$).

This gives some information about the relative sizes of the classes $C(v)$. Given a set of vertices X, let $N(X)$ be the total amount of elements in all $C(x)$, where $x \in X$. The objective of the paper is to bound $N(Y)/N(X)$ for given vertex sets X and Y.

In order to obtain bounds on $N(Y)/N(X)$, we write the constraints as a system of linear inequalities. Let $s'(v, w)$ be the (unknown) amount of switchings going from $C(v)$ to $C(w)$. For instance, for any vertex v, we have $\sum s'(vw) \geq a(v)N(v)$. A nonzero assignment of s and N satisfying the full set of inequalities is called a *feasible solution*, and a feasible solution which maximizes $N(Y)/N(X)$ is called an *optimal solution*.

The key result of the paper is to show that optimal solutions always take one of six standard forms. This is proved by a reduction of the inequalities to a linear program, and the fact that an optimal solution of a linear program always occurs at a vertex of the corresponding convex polytope. The analysis is amenable to a general bound α on the arcs of the digraph instead of the functions a, b (in that case we have $\alpha(v,w) = b(w)/a(v)$). Althoughα doesn't have a clear combinatorial interpretation, the paper suggests that different α may arise from richer information bounding the behaviour of the flow of switchings.

For a number of commonly-satisfied assumptions, the paper proves some alternative bounds for $N(Y)/N(X)$ that are slightly looser but easier to apply. In particular, a common use case is that we have some statistic on the objects in Ω and we believe that objects with a high value of that statistic make up a negligible fraction of all the objects in Ω. We would then partition the objects according to our statistic, and design a switching that tends to decrease the statistic, but by no more than some fixed amount. We would choose X to be the set of all vertices (partititions) and Y would be the set of all vertices (partitions) with a statistic value higher than some M.

1.2 Remarks

- I'm not sure why $i_{k-1} = \max\{M - (k-1)K, N+1\}$ is required for cases (a) and (b) of Lemma 3, instead of just $M - (k-1)K$ as in case (c).
- In Corollary 1, it isn't spelled out what assumptions X should satisfy, but it would seem A1 still has to hold.
- In Corollary 1, the somewhat inscrutable expression $k = \left\lceil \frac{M+\min\{0,K-\rho-1\}}{K} \right\rceil$ can be more easily seen to be $k = \left\lceil \frac{M-\lfloor\rho\rfloor}{K} \right\rceil$.

2 Asymptotic enumeration of sparse multigraphs with given degrees

C. Greenhill and B. D. McKay, Asymptotic enumeration of sparse multigraphs with given degrees, arXiv preprint arXiv:1303.4218 (2013)

2.1 Summary

The paper introduces a generalized version of the switching method, where there are multiple different "colours" of switchings that operate on the same set of objects. By slightly loosening the bound, this generalized model can be reduced to an instance of the single-switching model.

2.2 Remarks

- The switchings are described in terms of *oriented* edges. That makes sense because M is the sum of the degrees and therefore twice the number of edges.
- When giving a lower bound for the number of switchings, there are formulas like $[\ell_1]_3 - O(k_{\max}\ell_1^2)$. The reasoning for the $O(k_{\max}\ell_1^2)$ term is that it bounds the number of triples where two vertices are loops and the second of those is adjacent to the final vertex. This is an upper bound for the number of triples of loops which contain an edge.

3 Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$

B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, Combinatorica, 11 (1991) 369-382.

4 Random Graphs

S. Janson, T. Łuczak and A. Ruciński, Random Graphs (2000).

4.1 Remarks

4.1.1 Proposition 1.15

The “subsubsequence principle” means that taking n along any subsequence of $\mathbb{N}$, we can find a subsubsequence so that either M is bounded, or $N-M$ is bounded, or $M(N-M)/N \to \infty$. For, suppose none of these are true. Take a subsubsequence where M/N and $(N-M)/N$ both converge. They cannot both converge to zero; if for example $M/N \to L > 0$ then $M(N-M)/N = \Theta(N-M)$, contradiction.

In fact we can assume that if $M(N-M)/N \not\to \infty$ then M or $N-M$ is constant, which simplifies the analysis significantly. If M is constant, prove that $\Pr\big(\left|\Gamma_{M/N}\right| < M\big) \to p_M(M)/e^M$, where p_M is the first M terms of the taylor expansion of e^z. Also, I think the result of the central limit theorem is not enough for the estimate $\Pr\big(\left|\Gamma_{M/N}\right| < M\big) \to 1/2$, one needs to dig into the proof by convergence of characteristic function.

4.1.2 Theorem 5.4

The process can be viewed as either exploring a random graph or randomly generating a graph in an exotic edge order. Both are equivalent, but different viewpoints help understand different parts of the argument. The “kth step” means the kth saturation, not the kth generation. If the process is dead before step k, then X_k is not defined. The last paragraph on p109/first paragraph on p110 is confusing and contains some typos, here’s a little reprashing of bits and pieces:

The event that after the first k steps (saturations) there are fewer than $(c-1)k/2$ unsaturated vertices remaining and the process is alive is

$$H = \left\{ \text{process is alive for step } k \text{and} \sum_{i=1}^{k} X_i - k < \frac{(c-1)k}{2} \right\}.$$

Now, at step i, if m vertices have been touched there are $n-m$ vertices that can be touched, each with probability p, so $X_i \sim \mathrm{Bi}(n-m,p)$. For event H, no more than $(c+1)k_+/2$ vertices can have been touched up to step k, so $X_i \geq X_i^-$ where $X_i^- \sim \mathrm{Bi}(n - k_+(c+1)/2, p)$. Then $\sum_{i=1}^k X_i^- \sim \mathrm{Bi}\big(k\big(n - n^{2/3}(c+1)\big)/2, p\big)$ and $\mathbb{E}\Big[\sum_{i=1}^k X_i^-\Big] > k-1$ for sufficiently large n.

The bound for there to be two giant components uses $e^Q \geq 1+Q$ for $Q = -c/n$. I think formal justification for the for the $o(1)$ term in $\rho_- + o(1)$ is quite complicated. Let $Y_n \sim \mathrm{Bi}(n, c/n)$ and $Y \sim \mathrm{Po}(c)$. Since $Y_n \xrightarrow{d} Y$, for all ε there is N so that $Y_n \leq (1+\varepsilon)Y$ for $n > N$. The pgf of Y is $G : x \mapsto \exp\big(c\big(x^{1+\varepsilon} - 1\big)\big)$ so by the implicit function theorem, the solution for x of $G(x;\varepsilon) = x$ is continuous in ε and

$$\begin{aligned}
&\Pr(\text{branching process with } X_i \sim \mathrm{Bi}(n, c/n) \text{ dies after more than } k_- \text{ steps}) \\
&\leq \Pr(\text{branching process with } X_i \sim \mathrm{Po}(c) \text{ dies after more than } k_- \text{ steps}) + o(1)
\end{aligned}$$

because Pr(process dies at all) is (in the limit) the same for both cases. But since $k_- \to \infty$ the above probability is $o(1)$.

To choose an ordered pair of small vertices, first choose a small vertex arbitrarily (there are $n\rho(n,p)$ ways to do this on average), then either choose a small vertex in the same component (there are at most $k_- - 1$ ways to do this), or choose a small vertex in a different component (there are independently $n\rho(n - O(k_-), p)$ ways of doing this on average). Using a similar bounding argument as for $\rho(n,p)$ we have $\rho(n - O(k_-), p) \sim \rho(n,p)$.

4.1.3 Theorem 5.5

The reasoning behind the estimate $k!k^2$ is that each bad subgraph can be constructed by choosing a path then choosing two points on that path to glue the endpoints to. The first line of the displayed inequality is due to Markov's inequality, the second can be derived with $\left(\frac{2M}{n}\right)^{k+1} > \frac{n^{k+1}[M]_{k+1}}{[n^2/2-1]_{k+1}} > \frac{[n]_k[M]_{k+1}}{\left[\binom{n}{2}\right]_{k+1}}$ for $2 \leq k < n/2$.

4.1.4 Proof of Theorem 9.23

For the trace identity, prove by induction that $\left(A^k\right)_{ij}$ is the sum of all products $a_{i\alpha_1}a_{\alpha_1\alpha_2}\ldots a_{\alpha_{k-2}\alpha_{k-1}}a_{\alpha_{k-1}j}$.

5 Fundamentals of Stein's Method

N. Ross, Fundamentals of Stein's Method, Probability Surveys, 8 (2000) 210-293.

5.1 Summary

5.1.1 Overview

Stein's method is a technique for bounding the "distance" between distributions in some metric of the form $d(X,Z) = \sup_{h\in\mathcal{H}}|\mathbb{E}h(X) - \mathbb{E}h(Z)|$, where $\mathcal{H}$ is a family of functions. If $\mathcal{H} = \left\{1_{(-\infty,x]} : x \in \mathbb{R}\right\}$, then this measures the maximum distance between distribution functions (Kolmogorov metric). It's often possible to transfer bounds between different metrics, so it's worthwhile to define some auxilary metrics for which it's easier to apply Stein's method.

I think the motivations behind Stein's method are twofold. First, it gives a concrete bound of the kind not usually provided by asymptotic methods. Second, it can be used to give uniform convergence of probabilities, which is sometimes a necessary ingredient to asymptotic results where we allow things to vary (for example, proving something about $\mathbb{G}(n,p)$)

The idea of Stein's method is that there is some characterizing operator $\mathcal{A}$ of a random variable Z so that $\mathbb{E}\mathcal{A}f(X) = 0$ for all f in some manageable family $\mathcal{F}$, precisely when X has the distribution of Z. For example, if $Z \sim \mathcal{N}(0,1)$ then we can choose $\mathcal{A}$ as $f \mapsto (x \mapsto f'(x) - xf(x))$ and $\mathcal{F}$ as the set of bounded, continuously differentiable functions. If $Z \sim \mathrm{Po}(\lambda)$ then we can choose $\mathcal{A}$ as $f \mapsto (k \mapsto \lambda f(k+1) - kf(k))$ and $\mathcal{F}$ as the set of all bounded functions with bounded differences.

For all $h \in \mathcal{H}$, we can try to choose f_h to solve (or approximately solve) the functional equation $\mathcal{A}f_h = h - \mathbb{E}hZ$ (and adjust/refine $\mathcal{F}$ to include all such f_h). Then, $d(X,Z) \leq \sup_{f\in\mathcal{F}}|\mathbb{E}\mathcal{A}f(X)|$. By the choice of $\mathcal{A}$, we hope that this bound is relatively sharp. Also, we also hope that bounding $|\mathbb{E}\mathcal{A}f(X)|$ is easier than directly bounding $|\mathbb{E}h(X) - \mathbb{E}h(Z)|$.

5.1.2 Exchangeable Pairs for Poisson Approximation

A pair of random variables (X, X') is exchangable if it has the same distribution as (X', X). If (X, X') are exchangeable, we have $\mathbb{E}[1\{X = X' + 1\}f(X)] = \mathbb{E}[1\{X' = X + 1\}f(X')]$ so $\mathbb{E}[\Pr(X' = X - 1|X)f(X)] = \mathbb{E}[\Pr(X' = X + 1|X)f(X + 1)]$.

So, if we have $\Pr(X' = X - 1|X) \approx cX$ and $\Pr(X' = X + 1|X) \approx c\lambda$ for some c then $\mathbb{E}[\lambda f(X) - Xf(X)] \approx 0$ and X is approximately Poisson.

To choose appropriate X', we want to perturb X in some "reversible" way, so that X can also be symmetrically viewed as a perturbation of X'. Sometimes X can be viewed as some statistic on a space Ω of combinatorial objects with some distribution π, so we can obtain X' with a reversible Markov chain on Ω that has stationary distribution π.

For example, if X is a sum of indicator variables $X_1, \dots, X_n$, we could view X as the cardinality of a random subset of $\{1, \dots, n\}$. For the transitions of a Markov chain, we could randomly select some $i \in \{1, \dots, n\}$ and independently re-determine its membership. This sort of construction often naturally gives a suitable exchangeable pair. The event $\{X' = X - 1\}$ can be interpreted as a "death" and the event $\{X' = X + 1\}$ can be interpreted as an "immigration" in the Markov chain. The probability of a death is roughly proportional to the number alive, and the probability of an immigration is roughly proportional to the sum of the probabilities for each element to be alive in the stationary distribution, which is $\mathbb{E}[X] \approx \lambda$.

5.2 Remarks

- For the proof of proposition 1, the $C\varepsilon/2$ term comes from a $C \times \varepsilon$ triangular region of integration. The $d_W(W, Z)/\varepsilon$ term comes from the fact that $h_{x,\varepsilon}\varepsilon$ is 1-Lipschitz.
- For the proof of Lemma 3.4, as well as Cauchy-Schwarz Hölder's inequality $\mathbb{E}|X| \leq \left(\mathbb{E}X^2\right)^{1/2}$ is used.
- For the proof of Theorem 3.5, the fact $\sum_{i=1}^n \sum_{j \in N_i} X_i = \left(\sum_{i=1}^n X_i\right)^2$ is used.

6 Exchangeable Pairs and Poisson Approximation

S. Chatterjee, P. Diaconis, E. Meckes, Fundamentals of Stein's Method, Probability Surveys, 2 (2005) 64-106.

7 Exchangeable Pairs, Switchings and Random Regular Graphs

T. Johnson, Exchangeable Pairs, Switchings and Random Regular Graphs.